Black-box recognition of finite simple groups of Lie type by statistics of element orders

László Babai*
University of Chicago and
Rényi Institute, Budapest

William M. Kantor*
University of Oregon

Péter P. Pálfy**
Eötvös University, Budapest

Ákos Seress*
The Ohio State University

Abstract

Given a black-box group G isomorphic to some finite simple group of Lie type and the characteristic of G, we compute the standard name of G by a Monte Carlo algorithm. The running time is polynomial in the input length and in the time requirement for the group operations in G.

The algorithm chooses a relatively small number of (nearly) uniformly distributed random elements of G, and examines the divisibility of the orders of these elements by certain primitive prime divisors. We show that the divisibility statistics determine G, except that we cannot distinguish the groups $\mathrm{P}\Omega(2m+1,q)$ and $\mathrm{PSp}(2m,q)$ in this manner when q is odd and $m \geq 3$. These two groups can, however, be distinguished by using an algorithm of Altseimer and Borovik.

2000 Mathematics Subject Classification: Primary 20D06; Secondary: 20-04, 20P05, 68W30.

* This research was supported in part by the National Science Foundation.

** This research was supported in part by grant OTKA T29132, Hungary.

1 Introduction

There have been a number of recent algorithms for recognizing finite groups of Lie type. Some of these [13, 25, 26] take a matrix group $G = \langle S \rangle \leq \mathrm{GL}(d, q)$ as input, and decide by a polynomial-time one-sided Monte Carlo algorithm whether G is a classical group defined on the d-dimensional vector space over $\mathrm{GF}(q)$. (We refer to Section 2 for the definition of Monte Carlo algorithms.) Other approaches [8, 12, 14] go further. They recognize G constructively, which means that they provide procedures that express any given element of G in terms of S. However, these algorithms do not run in polynomial time if the field size q is not polynomial in the input length.

Still other approaches [17, 7, 9, 19, 20, 21] consider constructive recognition in the more general situation when a simple group $G = \langle S \rangle$ is given as a *black-box group*, where "constructive" means that they construct an isomorphism with a "concrete" copy of the group; but again the running time is not polynomial for large q. Recall that the elements of a black-box group G are assumed to be coded by 0-1 strings of uniform length N. A group element may be encoded by different strings and there may be strings which are not the coding of any group element. Oracles are provided for multiplying or inverting elements and for deciding whether or not two given elements are equal. In black-box groups, we automatically have the upper bound 2^N for $|G|$, and so $N \geq \log |G|$. For example, if G is a classical group of dimension d over $\mathrm{GF}(q)$ then $d^2 \log q$ is $O(N)$.

In this paper we also consider simple groups of Lie type given as black-box groups. Our goal is less ambitious than constructive recognition, but our algorithm runs in polynomial time (meaning $O(\mu |S| N^c)$ time, where μ is an upper bound for the time requirement of group operations in G and c is an absolute constant).

Theorem 1.1 *There is a polynomial-time Monte Carlo algorithm which, when given a black-box group $G = \langle S \rangle$ known to be isomorphic to a finite simple group of Lie type in given characteristic p, finds the standard name of G.*

The proof involves information concerning the proportions of elements of G of certain carefully chosen orders. This is similar in spirit to statistical ideas used in the aforementioned references or in [5]. We construct a sample of (nearly) uniformly distributed random elements of G, and determine whether the orders of these elements are divisible by certain primitive prime divisors (cf. Section 2). In Sections 3–4 we describe which primitive prime divisors enable us to distinguish the different groups of Lie type. Section 5 contains probability estimates that are used in Section 6 to deduce that sampling $O(N)$ elements provides the correct divisibility statistics with large probability. We note that if an upper bound $M < 2^N$ is known in advance for $|G|$ then a sample of size $O(\log M)$ suffices, but we formulate our results using only the bound $M = 2^N$.

Our method determines the standard name of G required in Theorem 1.1, except that a different and more delicate argument is required to distinguish the groups $\mathrm{P\Omega}(2m+1, q)$ and $\mathrm{PSp}(2m, q)$ when q is odd and $m \geq 3$ [1].

In contrast to other recent Monte Carlo recognition algorithms for classical groups [13, 25, 26] mentioned above, we do not even start with knowledge of the correct dimension or field, thereby enhancing the possibilities for applications of our results (e.g., in [5, 24]). As with other algorithmic investigations into groups of Lie type, not having linear algebra available has required entirely different types of methodologies to be developed. We have assumed that the characteristic of our group G is known in advance. That assumption can be avoided in various settings (cf. [5, 21]).

Theorem 1.1 proves Conjecture 9.2 in [5]. Portions of that theorem, proved by the first and third authors in [6], were first announced in [5]. After producing all of Table 1 except for the last column, they discovered that its entries v_1, v_2 had been used long ago by Artin [2] to help distinguish simple groups by their orders (note, however, that we do not know $|G|$ in Theorem 1.1). The invariants v_1, v_2 were also introduced by the remaining two authors in [20]. Upon receipt of [5], they realized that its Conjecture 9.2 could be proved using the methods of [21], which led to the present merged paper.

2 Background

A randomized algorithm is called *Monte Carlo* if it may return an incorrect output, but the probability of error is controlled by the user (see [4] for a discussion of randomized algorithms). A "one-sided" Monte Carlo algorithm for a decision problem means that one of the possible two outputs is guaranteed to be correct. In the context of recognition algorithms for classical groups of Lie type in their natural representation, this means that if the algorithm outputs that G is a classical group then the output is correct.

We refer the reader to [3, 4, 5, 20, 21] for discussions of black-box groups and previous algorithms for recognizing these groups. We emphasize again that, while the algorithms in [20, 19] constructively recognize the black-box groups in Theorem 1.1 (and hence provide more information than that theorem), those algorithms do not run in time polynomial in N when the size of the underlying field is not bounded.

The above references also discuss the role of black box groups in the study of groups of matrices over finite fields; this is the most important case of black-box groups. Here we only note that it is possible that the oracles for a black-box group perform the group operations in an overgroup $\overline{G}$ of G (the example we have in mind is $G \leq \overline{G} = \mathrm{GL}(V)$). In this case, we assume that the oracles can test whether a string represents an element of $\overline{G}$ (and so the group operations can be performed), but we do not assume that the oracles can decide whether a string represents an element of $\overline{G} \setminus G$ or G.

We will need random elements of black-box groups. We say that an algorithm outputs an *ε–uniformly distributed element* x in a group G if $(1-\varepsilon)/|G| < \mathrm{Prob}(x = g) < (1+\varepsilon)/|G|$ for all $g \in G$. *Nearly uniform* means ε–uniform for some $\varepsilon \leq 1/2$.

Theorem 2.1 [3] *Let c and C be given positive constants. Then there is a Monte Carlo algorithm which, when given a black-box group $G = \langle S \rangle$ of order at most M, sets up a data structure for the construction of ε-uniformly distributed elements for $\varepsilon = M^{-c}$, at a cost of $O(\log^5 M + |S| \log\log M)$ group operations. The probability that the algorithm fails is at most M^{-C}.*

If the algorithm succeeds, it permits the construction of ε–uniformly distributed, independent random elements of G at a cost of $O(\log M)$ group operations per element.

A fundamental and standard notion used throughout this paper is that of a *primitive prime divisor*. Let q be a prime power. An *odd* prime r is called a primitive prime divisor of $q^k - 1$, and called a $\mathrm{ppd}(q;k)$–prime, if $r \big| q^k - 1$ but $r \not| q^i - 1$ for $1 \le i < k$. Note that we do not allow 2 to be a primitive prime divisor. By a theorem of Zsigmondy [28], if p is prime then $\mathrm{ppd}(p;k)$–primes exist except when either $p = 2$, $k = 6$, or $p = 2$, $k = 1$, or $k = 2$ and p is a Mersenne prime, or $k = 1$ and p is a Fermat prime. These exceptions will require some extra work that will occur mainly in Section 4.

We will call an integer j a $\mathrm{ppd}^{\#}(q;k)$–*number* if j is divisible by a primitive prime divisor of $q^k - 1$. Furthermore, j is called $\mathrm{ppd}^{\#}(q;k_1) \cdot \mathrm{ppd}^{\#}(q;k_2)$–number if it is both $\mathrm{ppd}^{\#}(q;k_1)$ and $\mathrm{ppd}^{\#}(q;k_2)$. We say that a group element is a $\mathrm{ppd}^{\#}(q;k)$–*element* if its order is a $\mathrm{ppd}^{\#}(q;k)$–number. We also say that such elements have $\mathrm{ppd}^{\#}(q;k)$–*order*.

For an integer a and prime r, we denote by $(a)_r$ the largest power of r dividing a, and write $(a)_{r'} = a/(a)_r$.

Proposition 2.2 *Let G be a simple group of Lie type of characteristic p, and let $b \ge 6$ be the smallest integer such that all prime divisors of $|G|$ different from p divide some $p^j - 1$ with $j \le b$. Let $1 \ne g \in G$ and $S = \prod_{1 \le i \le b}(p^i - 1)$. Then*

(a) $b \le \lceil \log |G| \rceil$.

(b) The order of the Sylow p-subgroups of G is less than p^{b^2}.

(c) If $t \ne p$ is a prime and G contains an element of order t^a for some $a \ge 1$, then $t^a \big| p^j - 1$ for some $j \le b$.

(d) g is semisimple if and only if $g^S = 1$.

(e) Let $k \ge b/6$. Then the order of g is divisible by a $\mathrm{ppd}(p;k)$–prime $r > 5$ if and only if $g^K \ne 1$, where

$$K = p^{b^2}(S)_2(S)_3(S)_5 \prod_{1 \le i \le b, k \not| i} (p^i - 1) \prod_{1 \le i \le b, k | i} \frac{p^i - 1}{p^k - 1}.$$

(f) If $r \in \{2, 3, 5\}$ and $r \ne p$ then the order of g is divisible by r if and only if $g^{p^{b^2}(S)_{r'}} \ne 1$.

Proof. (a) and (b): These statements follow easily from the order formulas for simple groups of Lie type.

(c): Any element of order t^a is in a maximal torus of G. The structure of maximal tori is known for all groups of Lie type. The maximal tori for classical

groups are described in [11], while the maximal tori of all exceptional groups are collected in [21] from the literature. Using these lists, it is straightforward to check that the assertion holds.

(d): This follows directly from (c).

(e): Let $r > 5$ be a ppd$(p;k)$–prime. Then $r \big| p^i - 1$ if and only if $k \big| i$. In this case

$$\frac{p^i - 1}{p^k - 1} = p^{i-k} + \ldots + p^k + 1 \equiv \frac{i}{k} \pmod{r}.$$

Since $1 \leq i/k \leq b/k \leq 6$, $(p^i - 1)/(p^k - 1)$ is not divisible by r. Hence $r \not| K$, so if the order of g is divisible by r then we have $g^K \neq 1$. Conversely, assume that the order of g is not divisible by any ppd$(p;k)$–prime $r > 5$. Let the prime factorization of the order of g be $p^{a_0} t_1^{a_1} \cdots t_n^{a_n}$. Clearly, $p^{a_0} \big| p^{b^2}$ by (b). For each $t = t_i$ $(i = 1, \ldots, n)$ we have to show that $t^a = t_i^{a_i}$ divides K. It obviously holds for $t \in \{2, 3, 5\}$, so assume $t > 5$. Let j be the smallest integer such that $t^a \big| p^j - 1$. By (c), such $j \leq b$ exists. If $k \not| j$ then $p^j - 1$ is a factor of K. If $k \big| j$ then let d be the integer such that t is a ppd$(p;d)$–prime. If $d \not| k$ then $t \not| p^k - 1$ and so t^a divides the factor $(p^j - 1)/(p^k - 1)$ of K. Finally, if $d \big| k$ then $d < k$ since t is not a ppd$(p;k)$–prime, and $k = j$ since $(p^k - 1)_t = (p^j - 1)_t$ because $j/k \leq 6$ and $t > 6$, and j is the smallest integer such that $t^a \big| p^j - 1$. Since $(p^k - 1)_t > (p^d - 1)_t$, we must have $t \big| k$ and $(p^k - 1)_t = (p^{k/t} - 1)_t \cdot t$. Therefore, $(p^k - 1)_t \leq (p^{k/t} - 1)_t \cdot (p^{2k/t} - 1)_t$, and $(p^{k/t} - 1)(p^{2k/t} - 1)$ is a factor of K.

(f): This is trivial. □

We note that the same type of result can be proved without the restriction $k \geq b/6$, but would involve handling all primes up to b separately.

In Section 3 we shall define the invariant v_1. With a few exceptions, this is the largest k such that the group contains elements of ppd$^{\#}(p;k)$–order. The crude bound given in Proposition 2.2(a) says that $v_1 \leq N$ (where N is the black-box group parameter in Theorem 1.1). In our algorithm we shall make use of ppd$(p;k)$–primes only for values of k in the range $v_1 \geq k \geq v_1/6$. Note that, for any positive integer K, we can compute g^K using $O(\log K)$ group multiplications by repeated squaring. Hence, after the value of v_1 is known, checking whether a given $g \in G$ has ppd$^{\#}(p;k)$–order for some $v_1 \geq k \geq v_1/6$ can be done in polynomial time.

3 Numerical invariants

If we use the generic notation $G = L(q)$ for a finite simple group of Lie type (including the twisted types) over the finite field of size $q = p^e$, then, as in [22, p. 96], $|G|$ can be expressed in the form

$$|G| = \frac{1}{d} P_L(q);$$

here $d = (n, q-1)$ for $G = \mathrm{PSL}(n,q)$, $d = (n, q+1)$ for $G = \mathrm{PSU}(n,q)$, and $d \leq 4$ in all other cases, and P_L is a polynomial. Moreover, P_L can be expressed

as a product of factors of the form q, $q^i - 1$, $q^i + 1$ (for twisted types), and $q^8 + q^4 + 1$ (for ${}^3D_4(q)$). If $\Phi_k(x)$ denotes the kth cyclotomic polynomial, then we obtain a factorization of the form

$$|G| = \frac{1}{d} p^{eh} \prod_k \Phi_k(p)^{r_k} \tag{3.1}$$

for positive integers h, k, and r_k (cf. [22, p. 101]).

Notation: We denote the largest, second largest, and third largest k such that $\Phi_k(p)$ occurs in this factorization of the order of G by $v_1 > v_2 > v_3$, respectively, and call them and $w = v_1/(v_1 - v_2)$ the "invariants" of G.

The invariants v_i can be determined easily by inspecting the order formulas, and are given in Table 1. (Artin [2] computed the values of v_1 and v_2, together with other numerical invariants, for the simple groups known at the time. His work was completed in [22, p. 114].) The blank entries in rows $\mathrm{PSL}(2,q)$ and ${}^2B_2(q)$ of the table either do not exist (in the case $\mathrm{PSL}(2,p)$ for prime p) or depend on the arithmetic structure of e in a more complicated fashion, and these entries are not used by our algorithm. In cases ${}^2G_2(3)$, $G_2(2)$, ${}^2F_4(2), \mathrm{Sp}(4,2)$, we will assume that the input group is isomorphic to the *simple* group ${}^2G_2(3)'$, $G_2(2)'$, ${}^2F_4(2)', \mathrm{Sp}(4,2)'$, respectively, which has the same v_i values as the corresponding group listed in the table.

The following lemma connects the factors $\Phi_k(p)$ occurring in the factorization of $|G|$ to the $\mathrm{ppd}(p;k)$–primes dividing the orders of group elements.

Lemma 3.2 *Let p be a prime, G a simple group of Lie type of characteristic p of order given by* (3.1)*, and $k \geq 2$.*

(a) Assume that $k \neq 6$ if $p = 2$, and $k \neq 2$ if p is a Mersenne prime. Then $\Phi_k(p)$ is a factor in (3.1) *if and only if $|G|$ has elements of* $\mathrm{ppd}^{\#}(p;k)$*–order.*

(b) Assume that $p > 2$. Then $\Phi_2(p)$ is a factor in (3.1)*.*

(c) $\Phi_1(p)$ is always a factor in (3.1)*.*

Proof. (a): Suppose that $\Phi_k(p)$ is a factor in (3.1), and let r be a $\mathrm{ppd}(p;k)$–prime. Then $r \big| \Phi_k(p)$. We claim that $r \big| |G|$. This is clear if $r \not| d$, so suppose that $r \big| d$. Since, by definition, $r > 2$, we have to deal with the following cases: $\mathrm{PSL}(n,q)$ with $r \big| (n, q-1)$ and $n \geq 3$; $\mathrm{PSU}(n,q)$ with $r \big| (n, q+1)$ and $n \geq 3$; $E_6(q)$ with $r = 3 \big| q - 1$; and ${}^2E_6(q)$ with $r = 3 \big| q + 1$. In all these cases the polynomial $P_L(q)$ is divisible by $(q-1)^2$, respectively by $(q+1)^2$, and so $r \big| P_L(q)/d$.

Conversely, assume that a $\mathrm{ppd}(p;k)$–prime r (with $k \geq 2$) divides $|G|$. Suppose first that we are not in the case $G = {}^3D_4(q)$, $r \big| q^8 + q^4 + 1$. Then $r \big| q^i - 1$ or $r \big| q^i + 1$ for an appropriate factor of $P_L(q)$. Here $k \big| ei$, and $k \big| 2ei$ but $k \not| ei$ (as $r > 2$), respectively, hence $\Phi_k(p)$ is a factor of $p^{ei} - 1$ and $p^{ei} + 1$, respectively. Suppose now that $G = {}^3D_4(q)$ and $r \big| q^8 + q^4 + 1$. If $r = 3$ then $k = 2$ and $\Phi_2(p)$ is a factor in (3.1). If $r > 3$ then $k \big| 12e$ but $k \not| 4e$, since otherwise we would have $q^4 \equiv 1 \pmod{r}$ and hence $q^8 + q^4 + 1 \equiv 3 \pmod{r}$, a contradiction. So $\Phi_k(p)$ is a factor in (3.1) in this case as well.

(b) and (c): These statements can be checked by straightforward inspection of the order formulas. $\square$

G	v_1	v_2	v_3
$\mathrm{PSL}(d,q), d \geq 4$	ed	$e(d-1)$	$e(d-2)$
$\mathrm{PSL}(2,q)$	$2e$	e	
$\mathrm{PSL}(3,q)$	$3e$	$2e$	$\begin{cases} e & \text{if } 2 \nmid e \\ 3e/2 & \text{if } 2 \mid e \end{cases}$
$\begin{cases} \mathrm{PSp}(2m,q) \\ \mathrm{P\Omega}(2m+1,q) \end{cases} m \geq 4$	$2em$	$e(2m-2)$	$e(2m-4)$
$\mathrm{PSp}(4,q)$	$4e$	$2e$	$\begin{cases} e & \text{if } 3 \nmid e \\ 4e/3 & \text{if } 3 \mid e \end{cases}$
$\begin{cases} \mathrm{PSp}(6,q) \\ \mathrm{P\Omega}(7,q) \end{cases}$	$6e$	$4e$	$3e$
$\mathrm{P\Omega}^+(2m,q), m \geq 6$	$e(2m-2)$	$e(2m-4)$	$e(2m-6)$
$\mathrm{P\Omega}^+(8,q)$	$6e$	$4e$	$3e$
$\mathrm{P\Omega}^+(10,q)$	$8e$	$6e$	$5e$
$\mathrm{P\Omega}^-(2m,q), m \geq 3$	$2em$	$e(2m-2)$	$e(2m-4)$
$\mathrm{PSU}(2m+1,q), m \geq 3$	$2e(2m+1)$	$2e(2m-1)$	$2e(2m-3)$
$\mathrm{PSU}(3,q)$	$6e$	$2e$	$\begin{cases} e & \text{if } 5 \nmid e \\ 6e/5 & \text{if } 5 \mid e \end{cases}$
$\mathrm{PSU}(5,q)$	$10e$	$6e$	$4e$
$\mathrm{PSU}(2m,q), m \geq 5$	$2e(2m-1)$	$2e(2m-3)$	$2e(2m-5)$
$\mathrm{PSU}(6,q)$	$10e$	$6e$	$4e$
$\mathrm{PSU}(8,q)$	$14e$	$10e$	$8e$
${}^2B_2(q)$	$4e$	$\begin{cases} e & \text{if } 3 \nmid e \\ 4e/3 & \text{if } 3 \mid e \end{cases}$	
${}^2G_2(q)$	$6e$	$2e$	$\begin{cases} e & \text{if } 5 \nmid e \\ 6e/5 & \text{if } 5 \mid e \end{cases}$
$G_2(q)$	$6e$	$3e$	$2e$
${}^3D_4(q)$	$12e$	$6e$	$3e$
${}^2F_4(q)$	$12e$	$6e$	$4e$
$F_4(q)$	$12e$	$8e$	$6e$
$E_6(q)$	$12e$	$9e$	$8e$
${}^2E_6(q)$	$18e$	$12e$	$10e$
$E_7(q)$	$18e$	$14e$	$12e$
$E_8(q)$	$30e$	$24e$	$20e$

Table 1: The invariants v_i

Remark 3.3 One may try to extend the equivalence in Lemma 3.2(a) to all values of p, k by defining $\mathrm{ppd}^{\#}(p;2)$–numbers for Mersenne primes and $\mathrm{ppd}^{\#}(p;1)$–numbers for Fermat primes as the numbers divisible by 4. However, in the groups $\mathrm{PSL}(2,5)$ and ${}^2G_2(3^e)$ there are no elements of order 4; among the simple groups with elementary abelian Sylow 2-subgroups these are the only exceptions concerning Fermat and Mersenne primes.

w	G
2	$\mathrm{PSL}(2,q), \mathrm{PSp}(4,q), G_2(q), {}^2F_4(q), {}^3D_4(q)$
$m \geq 3$, integer	$\mathrm{PSL}(m,q), \mathrm{PSp}(2m,q), \Omega(2m+1,q),$ $\mathrm{P\Omega}^-(2m,q), \mathrm{P\Omega}^+(2m+2,q)$
3	$F_4(q), {}^2E_6(q)$ and those listed for $m \geq 3$, using $m = 3$
4	$E_6(q)$ and those listed for $m \geq 3$, using $m = 4$
5	$E_8(q)$ and those listed for $m \geq 3$, using $m = 5$
3/2	$\mathrm{PSU}(3,q), {}^2G_2(q), {}^2B_2(2^e)$ with $3 \mid e$
$m/2 \geq 5/2, m$ odd	$\mathrm{PSU}(m,q), \mathrm{PSU}(m+1,q)$
9/2	$E_7(q), \mathrm{PSU}(9,q), \mathrm{PSU}(10,q)$
4/3	${}^2B_2(2^e)$ with $3 \nmid e$

Table 2: The invariant $w = v_1/(v_1 - v_2)$

The situation is much worse in the case $p = 2$, $k = 6$. The natural definition is that $\mathrm{ppd}^{\#}(2;6)$–numbers are the numbers divisible by 9 and we have $\Phi_6(2) = 3$. However, the following groups of characteristic 2 have order divisible by 3 but do not contain an element of order 9: $\mathrm{PSL}(5,2^e)$ and $\mathrm{PSL}(4,2^e)$ with $(6,e) = 1$; $\mathrm{PSL}(3,2^e)$ and $\mathrm{PSL}(2,2^e)$ with $(3,e) = 1$; $\mathrm{PSU}(5,2^e)$ and $\mathrm{PSU}(4,2^e)$ with $(6,e) = 2$; $\mathrm{PSU}(3,2^e)$ with $(3,e) = 1$; $\mathrm{PSp}(4,2^e)$ with $(3,e) = 1$; ${}^2F_4(2^e)$ with $(6,e) = 1$; and $G_2(2^e)$ with $(3,e) = 1$. It is straightforward to compile this list (e. g., from [21] by examining the tables given in Section 2 of that paper for the exceptional groups, and using Propositions 3.2, 3.18, 3.28, 3.39 for the classical groups).

These difficulties make it somewhat awkward to extend the definition of $\mathrm{ppd}^{\#}(p;k)$–numbers to consider the aforementioned situations (cf. [20, Section 2.4]), but we will employ such an extension in Subsection 4.3.

Now we begin collecting the data for each group which enables us to distinguish it from the other groups. The first two items we consider are the values of v_1 and v_2. We classify the groups according to the invariant w in Table 2. As in [20, Section 7.2.1], it is easy to use Table 2 in order to check the following crucial fact:

Proposition 3.4 *There are at most seven groups with the same pair of invariants* (v_1, v_2).

Note that $\mathrm{PSU}(4,q) \cong \mathrm{P\Omega}^-(6,q)$ and, contrary the usual convention, we prefer to use the latter group in Tables 1 and 2 since it better fits our general pattern.

4 Distinguishing groups with the same (v_1, v_2)

In this section we will describe additional information for each group which distinguishes the various groups with the same pair of invariants (v_1, v_2). We use three different types of data items:

(a) for appropriately chosen y_1 and y_2 (and in one case y_3 as well) the information whether the group contains elements of $\mathrm{ppd}^{\#}(p;y_1)\cdot\mathrm{ppd}^{\#}(p;y_2)$–order;

(b) the value of v_3; and

(c) the information that the proportion of elements of a certain order is less than a specific bound c (or greater than a bound c) in the group.

We will show that such data distinguishes all pairs of groups with the same parameter pair (v_1,v_2), except $\mathrm{PSp}(2m,q)$ and $\Omega(2m+1,q)$ with q odd and $m\geq 3$. In the latter groups the order statistics are very similar, hence only a completely different method can distinguish these groups [1].

We begin by noting that the only case with odd v_1 occurs when $G\cong\mathrm{PSL}(m,p^e)$ with $v_1=me$ odd. Here $e=v_1-v_2$ and $m=v_1/e$, so (v_1,v_2) uniquely determines the isomorphism type of G. So henceforth *we may assume that v_1 is even.*

In the following three subsections, we describe the data which distinguishes classical groups from exceptional ones, exceptional groups among themselves, and classical groups among themselves, respectively. In each subsection, we organize our argument according to the w values defined in Table 2. In all cases we assume that the groups under consideration have the same invariants v_1 and v_2.

The information about tori in classical groups used in the argument can be found in [11], while the information about tori and orders of elements in exceptional groups is in the tables in [21, Section 2], where the original references are also provided.

4.1 Distinguishing classical groups from exceptional ones

$\underline{w=2}$

If $6\nmid v_1$ then the exceptional groups do not occur, so we may assume that $6|v_1$. Consider first the case $12|v_1$. Then both $\mathrm{PSL}(2,p^{v_1/2})$ and $\mathrm{PSp}(4,p^{v_1/4})$ contain elements of $\mathrm{ppd}^{\#}(p;v_1)\cdot\mathrm{ppd}^{\#}(p;v_1/3)$–order. On the other hand, ${}^3D_4(p^{v_1/12})$, ${}^2F_4(p^{v_1/12})$, and $G_2(p^{v_1/6})$ contain no such elements, since in these groups the maximal tori of $\mathrm{ppd}(p;v_1)$–order have order dividing $p^{v_1/3}-p^{v_1/6}+1$, where $(p^{v_1/3}-p^{v_1/6}+1,p^{v_1/3}-1)=(3,p^{v_1/6}+1)$. So the only possibility is that 3 is a $\mathrm{ppd}(p;v_1/3)$–prime, but this case is excluded since we assumed that $v_1\geq 12$.

If $v_1/6$ is odd then we have to distinguish only $\mathrm{PSL}(2,p^{v_1/2})$ and $G_2(p^{v_1/6})$. In the case $p=2$, $v_1=18$, the group $\mathrm{PSL}(2,2^9)$ contains elements of $9\cdot\mathrm{ppd}^{\#}(2;18)$–order, whereas $G_2(2^3)$ does not. In every other situation in which $v_1\geq 18$, observe that $\mathrm{PSL}(2,p^{v_1/2})$ contains elements of $\mathrm{ppd}^{\#}(p;v_1)\cdot\mathrm{ppd}^{\#}(p;v_1/3)$–order whereas $G_2(p^{v_1/6})$ does not.

Finally, suppose that $v_1=6$. If $p=2$ note that $\mathrm{PSL}(2,8)$ contains elements of order 9 while $G_2(2)'$ has none. Suppose that $p\geq 3$. Then $PSL(2,p^3)$ contains cyclic tori of order $(p^3+1)/2$, while in $G_2(p)$ all $\mathrm{ppd}^{\#}(p;6)$-orders of elements divide p^2-p+1. If $p\equiv 2 \pmod 3$, then 9 does not divide p^2-p+1, but

$(p^3+1)/2$ is divisible by 9, so we can distinguish the two groups by checking if they contain elements of $9 \cdot \mathrm{ppd}^{\#}(p;6)$-order. If p is a Mersenne prime, then $PSL(2,p^3)$ has elements of $2 \cdot \mathrm{ppd}^{\#}(p;6)$-order, while $G_2(p)$ does not. In all other cases, there exists a primitive prime divisor of p^2-1, moreover p^2-p+1 and $(p+1)/2$ are coprime, hence we can distinguish the two groups by checking if they contain elements of $\mathrm{ppd}^{\#}(p;6) \cdot \mathrm{ppd}^{\#}(p;2)$-order.

$\underline{w=3}$

The exceptional groups do not occur unless $12\big|v_1$ or $18\big|v_1$. If $18\big|v_1$ then the value of v_3 distinguishes ${}^2E_6(p^{v_1/18})$ from the other groups. (Note that in the case $p=2$, $v_1=18$, the third largest value of k for which elements of $\mathrm{ppd}^{\#}(2;k)$–order occur in $\mathrm{P\Omega}^-(6,8)$ is 4, so we do not detect $v_3=6$ using Lemma 3.2, but this does not influence the distinction of ${}^2E_6(2)$ from the other groups.)

In the next two paragraphs we will use the fact that the only maximal torus of $\mathrm{ppd}^{\#}(p;v_1)$–order in $F_4(p^{v_1/12})$ has order $p^{v_1/3}-p^{v_1/6}+1$.

If $12\big|v_1$ and $v_1>12$, or $v_1=12$ and $p>2$, then $\mathrm{P\Omega}^-(6,p^{v_1/6})$ is distinguished from $F_4(p^{v_1/12})$ by the value of v_3. The group $\mathrm{PSL}(3,p^{v_1/3})$ is distinguished from $F_4(p^{v_1/12})$ because it contains elements of $\mathrm{ppd}^{\#}(p;v_1)\cdot\mathrm{ppd}^{\#}(p;v_1/2)$–order; and $\mathrm{PSp}(6,p^{v_1/6})$, $\Omega(7,p^{v_1/6})$, $\mathrm{P\Omega}^+(8,p^{v_1/6})$ are distinguished from $F_4(p^{v_1/12})$ because they contain elements of $\mathrm{ppd}^{\#}(p;v_1)\cdot\mathrm{ppd}^{\#}(p;v_1/3)$–order.

Finally, in the case $v_1=12$, $p=2$ the groups $\mathrm{P\Omega}^-(6,4)$ and $\mathrm{PSL}(3,16)$ are distinguished from $F_4(2)$ because they do not contain elements of order divisible by 9 whereas $F_4(2)$ does; and $\mathrm{PSp}(6,4)\cong\Omega(7,4)$, $\mathrm{P\Omega}^+(8,4)$ are distinguished from $F_4(2)$ because they contain elements of order 65 (which is a $\mathrm{ppd}^{\#}(p;v_1)\cdot\mathrm{ppd}^{\#}(p;v_1/3)$–number) whereas $F_4(2)$ does not.

$\underline{w\in\{4,5,9/2\}}$

The value of v_3 distinguishes the exceptional groups from the classical ones.

$\underline{w=3/2}$

If $p=2$ then $\mathrm{PSU}(3,p^{v_1/6})$ contains elements of order 3 whereas ${}^2B_2(2^{v_1/4})$ does not. If $p=3$ then $\mathrm{PSU}(3,p^{v_1/6})$ has elements of order 4 whereas ${}^2G_2(3^{v_1/6})$ does not.

4.2 Distinguishing exceptional groups

The value of w determines the group uniquely in the cases $w\in\{4,5,9/2,4/3\}$. In the case $w=3/2$, the exceptional groups are defined in different characteristics. In the case $w=3$, the value of v_3 distinguishes ${}^2E_6(p^{v_1/18})$ from $F_4(p^{v_1/12})$.

The case $w=2$ requires slightly more work. Exceptional groups occur only when $6\big|v_1$; moreover, if $v_1/6$ is odd then only $G_2(p^{v_1/6})$ occurs. It remains to consider the case $12\big|v_1$. Here the value of v_3 distinguishes ${}^3D_4(p^{v_1/12})$ from the other two groups. (Note that in the case $p=2$, $v_1=24$, the third largest value of k for which elements of $\mathrm{ppd}^{\#}(2;k)$–order occur in ${}^3D_4(4)$ is 4, so we do not

detect $v_3 = 6$ using Lemma 3.2, but this does not influence the distinction from $G_2(16)$; and there is no group ${}^2F_4(q)$ with $v_1 = 24$.) Finally, $G_2(2^{v_1/6})$ and ${}^2F_4(2^{v_1/12})$ are distinguished because $G_2(2^{v_1/6})$ has elements of $\mathrm{ppd}^{\#}(2; v_1/4)$–order whereas ${}^2F_4(2^{v_1/12})$ does not.

4.3 Distinguishing classical groups

In this subsection only, we define $\mathrm{ppd}^{\#}(2;6)$–numbers as the numbers divisible by 9; for Mersenne primes p, we define $\mathrm{ppd}^{\#}(p;2)$–numbers as the numbers divisible by 4; and for Fermat primes $p > 3$, we define $\mathrm{ppd}^{\#}(p;1)$–numbers as the numbers divisible by 4. This terminology helps us state the results of the subsection more uniformly.

If $w > 3/2$ is not an integer then $w = m/2$ for an odd integer $m \geq 5$, and we have to distinguish $\mathrm{PSU}(m+1, p^e)$ from $\mathrm{PSU}(m, p^e)$.

Lemma 4.1 *Let $m \geq 5$ be odd.*

(a) If $4|m+1$ then $\mathrm{PSU}(m+1, p^e)$ contains elements of $\mathrm{ppd}^{\#}(p; (m+1)e)$–order whereas $\mathrm{PSU}(m, p^e)$ does not.

(b) If $4 \nmid m+1$ then $\mathrm{PSU}(m+1, p^e)$ contains elements of $\mathrm{ppd}^{\#}(p; (m+1)e/2)$–order whereas $\mathrm{PSU}(m, p^e)$ does not.

Proof. (a) The factorization (3.1) for $|\mathrm{PSU}(m+1, p^e)|$ contains the term $\Phi_{(m+1)e}(p)$ and so, by Lemma 3.2, $\mathrm{PSU}(m+1, p^e)$ contains elements of $\mathrm{ppd}^{\#}(p; (m+1)e)$–order. (Since $4|(m+1)e$, the case $p = 2, (m+1)e = 6$ cannot occur.) On the other hand, the factorization (3.1) for $|\mathrm{PSU}(m, p^e)|$ does not contain the term $\Phi_{(m+1)e}(p)$ and hence $\mathrm{PSU}(m, p^e)$ has no elements of $\mathrm{ppd}^{\#}(p; (m+1)e)$–order.

(b) Similarly, if $(m+1)/2$ is odd then the factorization (3.1) for $|\mathrm{PSU}(m+1, p^e)|$ contains the term $\Phi_{(m+1)e/2}(p)$, and $\mathrm{PSU}(m+1, p^e)$ contains elements of $\mathrm{ppd}^{\#}(p; (m+1)e/2)$–order either by Lemma 3.2 or, in the case $p = 2$, $(m+1)e/2 = 6$, by inspection of the group $\mathrm{PSU}(6, 4)$. On the other hand, the factorization (3.1) for $|\mathrm{PSU}(m, p^e)|$ does not contain the term $\Phi_{(m+1)e/2}(p)$ and hence $\mathrm{PSU}(m, p^e)$ has no elements of $\mathrm{ppd}^{\#}(p; (m+1)e/2)$–order by Lemma 3.2 or by the inspection of $\mathrm{PSU}(5, 4)$. □

For integer values of w, most cases are covered by the following lemma.

Lemma 4.2 *In Table 3, '+' indicates if an element of $\mathrm{ppd}^{\#}(p; y_1) \cdot \mathrm{ppd}^{\#}(p; y_2)$–order occurs in the group in column 1, and '−' if it does not.*

There are three exceptions: (a) $p = 2, m = 3, e = 2$; (b) $p = 2, m = 3, e = 1$; and (c) $p = 2, m = 6, e = 1$, in which cases $\mathrm{PSL}(m, p^{2e})$ has no elements of $\mathrm{ppd}^{\#}(p; 2me) \cdot \mathrm{ppd}^{\#}(p; me)$–order. In case (b), $\mathrm{PSL}(m, p^{2e})$ is distinguished from the other groups by the fact that it does not have elements of order 9 while the other groups do; in cases (a) and (c), $\mathrm{PSL}(m, p^{2e})$ has elements of order $7 \cdot 13$, while the other groups do not.

Proof. See Propositions 3.3, 3.19, 3.29, 3.40 of [21]. □

	$m \geq 3$	$m \geq 4$	$m \geq 5$	$m \geq 6$	$m \geq 3$
parity of m		even	odd	even	odd
y_1	$2me$	$(m+2)e$	$(m+3)e$	$(m+2)e$	$(m+1)e$
y_2	me	me	$(m-1)e$	$(m-2)e$	$(m-1)e$
$\mathrm{PSL}(m, p^{2e})$	$+$				
$\mathrm{P\Omega}^+(2m+2, p^e)$	$-$	$+$	$+$		
$\mathrm{PSp}(2m, p^e)$, $\Omega(2m+1, p^e)$	$-$	$-$	$-$	$+$	$+$
$\mathrm{P\Omega}^-(2m, p^e)$	$-$	$-$	$-$	$-$	$-$

Table 3: The integers y_i

Because of the restrictions on m we had to make in Lemma 4.2, we will need to use other methods to distinguish

(a) $\mathrm{PSL}(2, p^{2e})$ from $\mathrm{PSp}(4, p^e)$,

(b) $\mathrm{P\Omega}^+(8, p^e)$ from $\mathrm{PSp}(6, p^e)$, $\Omega(7, p^e)$, and $\mathrm{P\Omega}^-(6, p^e)$, and

(c) $\mathrm{PSp}(8, p^e)$ and $\Omega(9, p^e)$ from $\mathrm{P\Omega}^-(8, p^e)$.

Case (a)

For distinguishing $\mathrm{PSL}(2, p^{v_1/2})$ and $\mathrm{PSp}(4, p^{v_1/4})$, we use the probability of an element having $\mathrm{ppd}^{\#}(p^{v_1/4}; 4)$–order. (Note that we use here $p^{v_1/4}$ instead of p.) Namely, Theorem 5.7 of Niemeyer and Praeger [26] yields that this probability lies in the interval $[1/3, 1/2)$ for $\mathrm{PSL}(2, p^{v_1/2})$, and in $[1/5, 1/4)$ for $\mathrm{PSp}(4, p^{v_1/4})$.

Case (b)

The value of v_3 distinguishes $\mathrm{P\Omega}^+(8, p^{v_1/6})$ from $\mathrm{P\Omega}^-(6, p^{v_1/6})$. Note that this is true even in the case $p = 2$, $v_1 = 12$, when $v_3 = 6$ in $\mathrm{P\Omega}^+(8, 4)$, since $\mathrm{P\Omega}^+(8, 4)$ contains elements of order 9.

In order to distinguish $\mathrm{P\Omega}^+(8, p^{v_1/6})$ from $\mathrm{PSp}(6, p^{v_1/6})$ and $\Omega(7, p^{v_1/6})$ let $e = v_1/6$, $q = p^e$, and observe that if $q > 3$ then the first group contains elements of $\mathrm{ppd}^{\#}(p; 4e) \cdot \mathrm{ppd}^{\#}(p; 2e) \cdot \mathrm{ppd}^{\#}(p; e)$–order, while the other groups do not. (This is the only place where we need the product of three $\mathrm{ppd}^{\#}(p; y_i)$-numbers.) Indeed, $\mathrm{P\Omega}^+(8, q)$ contains a torus of order $(q^4 - 1)/(4, q^4 - 1)$; note that this number is divisible by 4 if q is odd, so our argument is valid in the cases when q is a Mersenne or Fermat prime as well. On the other hand, in $\mathrm{PSp}(6, q)$ and in $\Omega(7, q)$ an element of $\mathrm{ppd}^{\#}(p; 4e) \cdot \mathrm{ppd}^{\#}(p; 2e)$–order has two irreducible nondegenerate subspaces of dimensions 4 and 2, its order divides $\mathrm{lcm}(q^2+1, q+1) = (q^2+1)(q+1)/(2, q-1)$, and hence its order has no $\mathrm{ppd}(p; e)$–prime divisor. (Note that if $q > 3$ is a Fermat prime then $(q^2+1)(q+1)/(2, q-1)$ is not divisible by 4.)

The only remaining cases are $q = 2$ and $q = 3$, where we use probability information contained in [16]. For $q = 2$, observe that the probability that an element has order 15 is $1/5$ in $\mathrm{P\Omega}^+(8, 2)$, while it is only $1/15$ in $\mathrm{PSp}(6, 2) \cong$

$\Omega(7,2)$. For $q = 3$, the probability that an element has order 20 is $3/20$ in $\mathrm{P}\Omega^+(8,3)$, while it is only $1/20$ in $\mathrm{PSp}(6,3)$ and $\Omega(7,3)$.

Case (c)

By [26, Theorem 5.7], the probability that a random element has $\mathrm{ppd}^{\#}(p^{v_1/8};8)$–order lies in the interval $[2/10, 2/8)$ for $\mathrm{P}\Omega^-(8,p^{v_1/8})$, but in $[1/9, 1/8)$ for $\mathrm{PSp}(8,p^{v_1/8})$ and $\Omega(9,p^{v_1/8})$.

5 Probability estimates

In Section 4, we described an assortment of integers y such that the existence or non-existence of elements of $\mathrm{ppd}^{\#}(p;y)$–, $2\cdot\mathrm{ppd}^{\#}(p;y)$–, $4\cdot\mathrm{ppd}^{\#}(p;y)$–, $9\cdot\mathrm{ppd}^{\#}(p;y)$–, $\mathrm{ppd}^{\#}(p;y_1)\cdot\mathrm{ppd}^{\#}(p;y_2)$– and $\mathrm{ppd}^{\#}(p;y_1)\cdot\mathrm{ppd}^{\#}(p;y_2)\cdot\mathrm{ppd}^{\#}(p;y_3)$–order in the input group G determines the isomorphism type of G. Some additional numbers of the same kind will be added to this list in Section 6, where we will also describe how to compute the value of v_1 and v_2. Algorithmically, we will decide whether G has elements of the required order by checking whether such orders occur in a random sample of elements. In this section, we give lower estimates for the proportion of the required element orders in G, which enable us to compute how many elements need to be sampled.

With one exception, the proportions of required elements in classical groups are covered by the following result (cf. Tables 1 and 3):

Theorem 5.1 [21, Theorem 5.6] *Let G be one of the simple classical groups defined on a vector space of dimension d, over the field $\mathrm{GF}(p^e)$.*

(1) *If r^a is a power of a prime $r \neq p$ such that $|G|$ has elements of order r^a, then there are at least $|G|/6d^2$ elements of G of order divisible by r^a.*

(2) *Assume that r^a and s^b are powers of distinct primes r, s such that G has an element of order $r^a s^b$ and, for some positive integers I, J, $r^a | p^{eI} - 1$ but $r^a \nmid p^{ek} - 1$ for $1 \leq k < I$ and $s^b | p^{eJ} - 1$ but $s^b \nmid p^{ek} - 1$ for $1 \leq k < J$. If $I + J \geq d - 1$ then there are at least $|G|/12d^2$ elements of S of order divisible by $r^a s^b$.*

The only case we will require that is not covered by Theorem 5.1 is the proportion of elements of $\mathrm{ppd}^{\#}(p;4e)\cdot\mathrm{ppd}^{\#}(p;2e)\cdot\mathrm{ppd}^{\#}(p;e)$–order in $\mathrm{P}\Omega^+(8,p^e)$ (cf. Subsection 4.3, Case (b)). *In Lemma* 5.2 *only*, similarly to Subsection 4.3, we define $\mathrm{ppd}^{\#}(2;6)$–numbers as the numbers divisible by 9; for Mersenne primes p, we define $\mathrm{ppd}^{\#}(p;2)$–numbers as the numbers divisible by 4; and for Fermat primes $p > 3$, we define $\mathrm{ppd}^{\#}(p;1)$–numbers as the numbers divisible by 4.

Lemma 5.2 *For $p^e > 3$ the proportion of elements of $\mathrm{ppd}^{\#}(p;4e)\cdot\mathrm{ppd}^{\#}(p;2e)\cdot\mathrm{ppd}^{\#}(p;e)$–order in $\mathrm{P}\Omega^+(8,p^e)$ is at least $1/60$.*

Proof. (See [21] for similar arguments used to prove Theorem 5.1.) It suffices to prove the same estimate for $G = \Omega^+(8,q) = \Omega^+(V)$. Decompose $V = W_2^+ \perp W_2^- \perp W_4^-$ with W_k^ε a nonsingular subspace of dimension k and type ε. There is an isometry group $X = A \times B \times C$ of V such that A, B, C induce cyclic groups of order $q-1$, $q+1$, q^2+1 on W_2^+, W_2^-, W_4^-, respectively, and the identity on the remaining two summands. Note that $X \cap G$ contains elements of the desired order, each of which uniquely determines the subspaces W_2^+, W_2^-, W_4^-. Then $C_G(X \cap G) = X \cap G$, $|N_G(X \cap G) \colon C_G(X \cap G)| = 2 \cdot 2 \cdot 4$, and the number of elements of the desired order in the union of all G-conjugates of $X \cap G$ is at least $|G \colon N_G(X \cap G)||X \cap G|(1-1/2)(1-1/3)(1-1/5) = |G|/60$. $\square$

The preceding results are needed in order to handle elements of G of order divisible by more than one prime of a suitable sort. When only one prime is involved, we can appeal to a much more general result:

Theorem 5.3 [18, Theorem 5.1] *Let G be a group of Lie type of characteristic p, and let h denote the Coxeter number of the Weyl group of the corresponding algebraic group. If $r \neq p$ is a prime divisor of $|G|$, then the probability is at least $(1-1/r)/h$ that an element of G has order divisible by r, except possibly when $r = 3$, G is $\mathrm{PSL}(3,q)$ or $\mathrm{PSU}(3,q)$, and this probability is at least $1/9$.*

Here the *Coxeter number* is the order of a Coxeter element of the group, and is as follows for the various types of groups [10, pp. 155,168]:

G:	A_l	B_l	C_l	D_l	G_2	F_4	E_6	E_7	E_8
h:	$l+1$	$2l$	$2l$	$2l-2$	6	12	12	18	30

.

In particular, if G is an exceptional group of Lie type then the stated probability is at least $(1/2)/30$ for *any* prime r other than the underlying characteristic. On the other hand, for classical groups we see that $h \leq d$, and hence the estimate in Theorem 5.3 is much better than the one in Theorem 5.1 for elements of order divisible by a prime, as opposed to a prime power or a product of primes.

The probability estimates in exceptional groups not covered by Theorem 5.3 are handled in the following lemma. These estimates will be needed in Section 6, for the computation of v_1 and v_2.

Lemma 5.4 *In each of the following cases, the proportion of elements with the described order is at least $2/21$: order divisible by 9 in $F_4(2)$; order divisible by 9 in ${}^3D_4(2)$; order 15 in $G_2(4)$; order 21 in $G_2(4)$.*

Proof. See [16]. $\square$

Finally, we describe an estimate which can be used to distinguish two groups by the proportions of elements of certain orders, when both groups contain such elements (cf. Subsection 4.3, Cases (a),(b),(c)). The method is based on Chernoff's bound [15]. Let $Y_1, \ldots, Y_t$ be not necessarily independent, $0,1$ valued random variables with the property that, for some r and each i, the conditional

probability $\text{Prob}(Y_i = 1 \mid Y_1 = x_1, \ldots, Y_{i-1} = x_{i-1}) \geq r$ for all 0-1 sequences $(x_1, \ldots, x_{i-1})$. Then, whenever $0 < \delta < 1$,

$$\text{Prob}\Big(\sum_{i=1}^{t} Y_i \leq (1-\delta)rt\Big) \leq e^{-\delta^2 rt/2}. \tag{5.5}$$

Lemma 5.6 *Suppose that the proportion of elements satisfying a certain property $\mathcal{P}$ is at most c_1 in group G_1 is at least c_2 in group G_2, for positive constants $c_1 < c_2$. Let $\varepsilon > 0$. If a given group G is isomorphic to G_1 or G_2 then, with probability greater than $1 - \varepsilon$, it can be determined which of G_1, G_2 our group G is isomorphic to by computing the proportion of elements satisfying $\mathcal{P}$ in a random sample of size $\lceil \ln(1/\varepsilon) \cdot \max\{8c_2/(c_2 - c_1)^2, 8(1 - c_1)/(c_2 - c_1)^2\} \rceil$.*

Proof. Take random elements $g_1, \ldots, g_t$ in G and define the $0, 1$ valued random variables Y_i by the rule that $Y_i = 1$ if and only if g_i has property $\mathcal{P}$. Let $X_i = 1 - Y_i$.

If $G \cong G_2$ then applying (5.5) with the parameters $r = c_2$, $\delta = (c_2 - c_1)/2c_2$, we obtain

$$\text{Prob}\Big(\sum_{i=1}^{t} Y_i \leq \frac{c_1 + c_2}{2} t\Big) \leq e^{-\frac{(c_2 - c_1)^2}{8c_2} t}.$$

On the other hand, if $G \cong G_1$ observe that $\text{Prob}(\sum_{i=1}^{t} Y_i \geq t(c_1 + c_2)/2) = \text{Prob}(\sum_{i=1}^{t} X_i \leq t(1 - (c_1 + c_2)/2))$. Applying (5.5) with the parameters $r = 1 - c_1$, $\delta = (c_2 - c_1)/2(1 - c_1)$, we obtain

$$\text{Prob}\Big(\sum_{i=1}^{t} Y_i \geq \frac{c_1 + c_2}{2} t\Big) \leq e^{-\frac{(c_2 - c_1)^2}{8(1 - c_1)} t}.$$

Therefore, choosing $t := \lceil \ln(1/\varepsilon) \cdot \max\{8c_2/(c_2 - c_1)^2, 8(1 - c_1)/(c_2 - c_1)^2\} \rceil$, and declaring that $G \cong G_1$ if $\sum_{i=1}^{t} Y_i \leq t(c_1 + c_2)/2$, the probability of error is less than ε. $\square$

6 An algorithm for Theorem 1.1

Given a simple group G of Lie type and its characteristic p, in this section we describe an algorithm that computes the standard name of G. Recall that N denotes the length of the 0-1 strings in the black-box group encoding of G.

Our first goal will be to find the value of v_1 for the input group G, based on the following lemma. For $g \in G$, let $h := g^{p^{N^2}}$ and define $j(g)$ to be the smallest nonnegative integer j such that $h^{\prod_{i=1}^{j}(p^i - 1)} = 1$. Note that, for any given $g \in G$, the value of $j(g)$ can be computed in polynomial time.

Lemma 6.1 *Let G be a simple group of Lie type of characteristic p, and let $v_1^* = \max_{g \in G} j(g)$. If $G \ncong \text{PSL}(6,2)$, $\text{PSL}(3,4)$, $\text{PSL}(2,8)$, $\text{PSp}(6,2)$, $\text{P}\Omega^-(6,2)$, $\text{P}\Omega^+(8,2)$ and $G_2(2)'$, then $v_1 = v_1^*$.*

Proof. By parts (a) and (b) of Proposition 2.2, $h = g^{p^{N^2}}$ has trivial p-part for any $g \in G$.

The exceptions listed in the statement of the lemma are the groups in characteristic 2 with $v_1 = 6$ (cf. the first column of Table 1). Hence Lemma 3.2(a) yields that v_1 is the largest integer with the property that $|G|$ is divisible by a $\mathrm{ppd}(p; v_1)$–prime, or $G \cong \mathrm{PSL}(2, p)$ with p Mersenne. If the order of some $g \in G$ is divisible by a $\mathrm{ppd}(p; v_1)$–prime then obviously $h^{\prod_{i=1}^{v_1-1}(p^i-1)} \neq 1$ and so $v_1^* \geq v_1$. Similarly, if $G \cong \mathrm{PSL}(2, p)$ with $p > 3$ Mersenne and the order of some $g \in G$ is divisible by 4 then $h^{p-1} \neq 1$ and so $v_1^* \geq v_1$. Conversely, Proposition 2.2(c) implies that for all $g \in G$ we have $h^{\prod_{i=1}^{v_1}(p^i-1)} = 1$ and so $v_1^* \leq v_1$. $\square$

We start the algorithm by computing the value of v_1^*. Given an arbitrary error bound ε, where $0 < \varepsilon < 1$, let $\mathcal{S}$ be a sample of group elements of size $\lceil \max\{24N \ln(1/\varepsilon), 60 \ln(1/\varepsilon)\} \rceil$. We claim that, with probability greater than $1 - \varepsilon$, if G is not $\mathrm{PSL}(2, p)$ with p Mersenne then $\mathcal{S}$ contains elements of $\mathrm{ppd}^{\#}(p; v_1^*)$–order, while if $G \cong \mathrm{PSL}(2, p)$ with $p > 3$ Mersenne then $\mathcal{S}$ contains elements of order divisible by 4. Indeed, if G is a classical group defined on a vector space of dimension d then $2^N > |G| > 2^{d^2/4}$ and, by Theorem 5.1(1), the probability that none of $\lceil 24N \ln(1/\varepsilon) \rceil$ random elements $g \in G$ have $j(g) = v_1^*$ is at most $(1 - 1/6d^2)^{24N \ln(1/\varepsilon)} < \varepsilon$. Similarly, if G is exceptional then, by Theorem 5.3, the probability that none of $\lceil 60 \ln(1/\varepsilon) \rceil$ random elements $g \in G$ have $j(g) = v_1^*$ is less than ε.

Similar probability estimates hold at every further step of the algorithm: at each step, either we apply Lemma 5.6 or we will have to decide whether G has elements of order divisible by a prime power or a product of two or three prime powers. (The only prime powers of exponent greater than one which occur in this context are 4 and 9.) If the answer is "yes" then the estimates in Theorems 5.1 and 5.3, and Lemmas 5.2 and 5.4 imply that a sample of size $\lceil \max\{24N \ln(1/\varepsilon), 60 \ln(1/\varepsilon)\} \rceil$ contains such elements with probability greater than $1 - \varepsilon$. Therefore, in the description of further steps we simply say "compute whether G has elements with a certain property", with the understanding that this can be done by sampling $O(N \log(1/\varepsilon))$ random elements. Since for any input group it is not hard to check that the number of steps of the algorithm is less than 15, the total number of random elements to be sampled is $O(N \log(1/\varepsilon))$.

Now we continue the description of the algorithm. If $p = 2$ and $v_1^* < 6$ then we compute whether G has elements of order divisible by 9; if the answer is "yes" then we replace v_1^* by 6. After that step $v_1^* = v_1$ for all simple groups, with two exceptions: in $\mathrm{PSL}(3, 4)$, $v_1 = 6$ and $v_1^* = 4$ and in $G_2(2)'$, $v_1 = 6$ and $v_1^* = 3$ (cf. Remark 3.3 and [16]).

Next, we determine the isomorphism type of G if $v_1^* \leq 4$, using the information in the first column of Table 1. If $v_1^* = 2$ then $G \cong \mathrm{PSL}(2, p)$. If $v_1^* = 3$ and $p > 2$ then $G \cong \mathrm{PSL}(3, p)$. If $v_1^* = 3$ and $p = 2$ then $G \cong PSL(3, 2)$ or $G \cong G_2(2)'$. The first of these has no elements of order 8, while the proportion

of elements of order 8 in $G_2(2)'$ is 1/4 [16]. Hence we can compute which one of these groups is (isomorphic to) G. If $v_1^* = 4$ then we compute whether G has elements of $\mathrm{ppd}^{\#}(p;3)$–order. If not, then $G \cong \mathrm{PSL}(2,p^2)$ or $\mathrm{PSp}(4,p)$ ($\mathrm{PSp}(4,2)'$ in the case $p=2$), and we can compute which one of these is G by the method described in Subsection 4.3, Case (a) and Lemma 5.6. If G contains elements of $\mathrm{ppd}^{\#}(p;3)$–order then, for $p>2$, we have $G \cong \mathrm{PSL}(4,p)$, and for $p=2$ we have $G \cong \mathrm{PSL}(4,2)$ or $\mathrm{PSL}(3,4)$. We can decide between the latter two groups since $\mathrm{PSL}(4,2)$ has elements of order 15 while $\mathrm{PSL}(3,4)$ does not [16].

Hence, from now on, we may assume that $v_1 = v_1^* \geq 5$ and we know the correct value of v_1. Our next goal is to find v_2. To this end, we compute

$$v_2^* := \max\{\, k < v_1 \mid \text{there exists a } \mathrm{ppd}^{\#}(p;k)\text{–element in } G \,\}.$$

The only groups for which v_2^* does not exist are ${}^2G_2(3)'$ and $\mathrm{PSU}(3,3)$ (cf. the second column of Table 1); we can decide between these two by the fact that $\mathrm{PSU}(3,3)$ contains elements of order 4 and ${}^2G_2(3)'$ does not [16]. Hence we may assume that v_2^* exists.

If p is odd then the only case when $v_2 \neq v_2^*$ is $G \cong \mathrm{PSU}(3,p)$ with $p>3$ Mersenne (cf. Table 1 and Lemma 3.2(a)), and this case can be recognized since this is the only one with $v_2^* = 1$. Thus, for odd p, we now know the value of v_2.

If $p=2$ and $v_2^* < 6 < v_1$ then we compute whether G has elements of order divisible by 9; if the answer is "yes" then we replace v_2^* by 6. After that step, $v_2^* = v_2$ for all simple groups in characteristic 2 as well, with two exceptions: in both ${}^2F_4(2)'$ and $G_2(4)$ we have $v_2 = 6$ and $v_2^* = 4$ (cf. Table 1 and Remark 3.3). In both of these groups, $v_1 = 12$. Hence, to finish the determination of v_2, we have to distinguish the groups with parameters $v_1 = 12$, $v_2^* = 4$; these groups are $\mathrm{PSU}(3,4)$, ${}^2B_2(8)$, ${}^2F_4(2)'$, and $G_2(4)$ by Table 1. We compute whether G has elements of order 3, 15, and 21. Then ${}^2B_2(8)$ is distinguished from the others as the only one with no elements of order 3; $G_2(4)$ is distinguished from the others as the only one with elements of order 21; and $\mathrm{PSU}(3,4)$ is distinguished from ${}^2F_4(2)'$ because it has elements of order 15 whereas ${}^2F_4(2)'$ has none [16].

Hence, for any input group, we know the value of v_1 and v_2 with large probability, and we can proceed to determine the standard name of G by computing the information described in Section 4. This finishes the description of the algorithm, and the proof of Theorem 1.1.

The algorithm was implemented by G. Malle and E. O'Brien, for matrix group inputs. The implementation follows quite closely this paper, with the exception of one subroutine. Instead of using Proposition 2.2(e) to check whether a given element $g \in G$ has $\mathrm{ppd}^{\#}(p;k)$–order, they compute the order of g, factorize the order, and decide whether a $\mathrm{ppd}^{\#}(p;k)$–prime occurs in this factorization.

If the algorithm is used for an input which is given as a factor group of a matrix group, or in any other situation where the order of group elements is not easily computable, then there is another way to avoid the time-consuming application of Proposition 2.2(e). This proposition gives a necessary and sufficient condition for a given element to have $\mathrm{ppd}^{\#}(p;k)$–order, hence allowing

the determination of the exact proportion of $\mathrm{ppd}^{\#}(p;k)$–orders in any sample of group elements. However, in most applications we do not need this exact proportion; we only have to establish that a group has $\mathrm{ppd}^{\#}(p;k)$–elements. In these cases, it is more efficient to apply a different, faster criterion which gives only a sufficient condition for a group element to have $\mathrm{ppd}^{\#}(p;k)$–order. Such criterion detects only the *subsets* of $\mathrm{ppd}^{\#}(p;k)$–elements which are used in the proofs of Theorems 5.1 and 5.3 and Lemma 5.2. Elements of these subsets are easier to identify algorithmically, but nevertheless we have seen that the subsets are large enough that random elements have a sufficiently large probability to belong to them.

Acknowledgement We are indebted to Gunther Malle and Eamonn O'Brien for pointing out an error in Subsection 4.1.

References

[1] C. Altseimer and A. V. Borovik. Probabilistic recognition of orthogonal and symplectic groups. In *Groups and Computation III* (eds. W. M. Kantor and Á. Seress), The Ohio State Univ. Math. Res. Inst. Publ. 8 (Walter deGruyter, Berlin–New York 2001), pp. 1–20.

[2] E. Artin. The orders of the classical simple groups. *Commun. Pure Appl. Math.* 8 (1955), 455–472.

[3] L. Babai. Local expansion of vertex-transitive graphs and random generation in finite groups. In *Proc. 23rd ACM Symp. on Theory of Computing* (1991), pp. 164–174.

[4] L. Babai. Randomization in group algorithms: conceptual questions. In *Groups and Computation II* (eds. L. Finkelstein and W. M. Kantor), DIMACS Series in Discrete Math. and Theoretical Computer Science 28 (AMS 1997), pp. 1–17.

[5] L. Babai and R. Beals. A polynomial-time theory of black-box groups I. In *Groups St Andrews 1997 in Bath, I* (eds. C. M. Campbell, E. F. Robertson, N. Ruskuc, and G. C. Smith), LMS Lecture Note Series 260 (Cambridge U. Press 1999), pp. 30–64.

[6] L. Babai and P. P. Pálfy. Recognizing finite simple groups by order statistics (manuscript, 1998).

[7] S. Bratus. Recognition of finite black-box groups. Ph.D. Thesis Northeastern U. (1999).

[8] P. A. Brooksbank. A constructive recognition algorithm for the matrix group $\Omega(d,q)$. In *Groups and Computation III* (eds. W. M. Kantor and Á. Seress), The Ohio State Univ. Math. Res. Inst. Publ. 8 (Walter deGruyter, Berlin–New York 2001), pp. 79–93.

[9] P. A. Brooksbank and W. M. Kantor. On constructive recognition of a black-box PSL(d, q). In *Groups and Computation III* (eds. W. M. Kantor and Á. Seress), The Ohio State Univ. Math. Res. Inst. Publ. 8 (Walter deGruyter, Berlin–New York 2001), pp. 95–111.

[10] R. W. Carter. *Simple groups of Lie type* (Wiley London-New York-Sydney-Toronto 1972).

[11] R. W. Carter. Centralizers of semisimple elements in the finite classical groups. *Proc. London Math. Soc.* 42 (1981), 1–41.

[12] F. Celler. Matrixgruppenalgorithmen in GAP. Ph. D. thesis. RWTH Aachen (1997).

[13] F. Celler and C. R. Leedham-Green. A non-constructive recognition algorithm for the special linear and other classical groups. In *Groups and Computation II* (eds. L. Finkelstein and W. M. Kantor), DIMACS Series in Discrete Math. and Theoretical Computer Science 28 (AMS 1997), pp. 61–67.

[14] F. Celler and C. R. Leedham-Green. A constructive recognition algorithm for the special linear group. In *The atlas of finite groups: ten years on* (eds. R. T. Curtis and R. A. Wilson), LMS Lecture Note Series 249 (Cambridge U. Press 1998), pp. 11–26.

[15] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statistics* 23 (1952), 493–507.

[16] J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker and R. A. Wilson. *Atlas of finite groups* (Clarendon Press, Oxford 1985).

[17] G. Cooperman, L. Finkelstein and S. Linton. Recognizing $GL_n(2)$ in non-standard representation. In *Groups and Computation II* (eds. L. Finkelstein and W. M. Kantor), DIMACS Series in Discrete Math. and Theoretical Computer Science 28 (AMS 1997), pp. 85–100.

[18] I. M. Isaacs, W. M. Kantor and N. Spaltenstein. On the probability that a group element is p–singular. *J. Algebra* 176 (1995), 139–181.

[19] W. M. Kantor and K. Magaard. Black-box exceptional groups of Lie type (in preparation).

[20] W. M. Kantor and Á. Seress. Black box classical groups. *Memoirs of the Amer. Math. Soc.* 149 (2001), No. 708.

[21] W. M. Kantor and Á. Seress. Prime power graphs for groups of Lie type. *J. Algebra*, to appear.

[22] W. Kimmerle, R. Lyons, R. Sandling and D. N. Teague. Composition factors from the group ring and Artin's theorem on orders of simple groups. *Proc. London Math. Soc.* (3) 60 (1990), 89–122.

[23] P. B. Kleidman and M. W. Liebeck. *The subgroup structure of the finite classical groups*. LMS Lecture Note Series 129 (Cambridge U. Press 1990).

[24] C. R. Leedham-Green. The computational matrix group project. In *Groups and Computation III* (eds. W. M. Kantor and Á. Seress), The Ohio State Univ. Math. Res. Inst. Publ. 8 (Walter deGruyter, Berlin–New York 2001), pp. 229–247.

[25] P. M. Neumann and C. E. Praeger. A recognition algorithm for special linear groups. *Proc. London Math. Soc.* (3) 65 (1992), 555–603.

[26] A. C. Niemeyer and C. E. Praeger. A recognition algorithm for classical groups over finite fields. *Proc. London Math. Soc.* (3) 77 (1998), 117–169.

[27] T. A. Springer and R. Steinberg. Conjugacy classes. In *A. Borel et al, Seminar on algebraic groups and related finite groups*, Lecture Notes 131 (Springer 1970), pp. E1-E100.

[28] K. Zsigmondy. Zur Theorie der Potenzreste. *Monatsh. Math. Phys.* 3 (1892), 265-284.